Scuola Superiore Sant'Anna
di Studi Universitari
e di Perfezionamento

Consorzio Nazionale
Interuniversitario
per le Telecomunicazioni

IMPLEMENTAZIONE DI UN SISTEMA DI PROTEZIONE GMPLS CONDIVISO PER RETI ETHERNET OTTICHE

Filippo Cugini, Luca Valcarenghi,
Piero Castoldi, Pier Giorgio Raponi

May, 2005

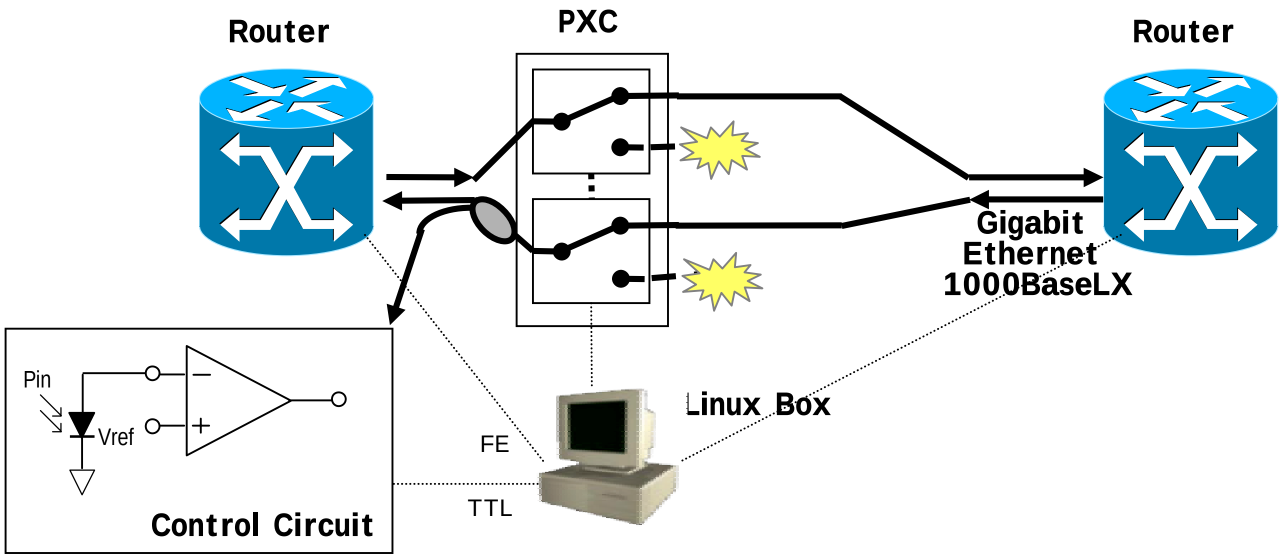
Introduction

- **Routed Optical Ethernet** networks (i.e., 1 and 10 GbE point-to-point connection between IP/MPLS routers) represent an appealing network solution for MAN because of its low cost and simplicity.
- A limiting factor for the full deployment of the Optical Ethernet architecture in MAN is the lack of some OA&M features, such as **efficient fault detection** and **recovery**.

Here we present:

- an experimental evaluation of **two** main limitations that affect failure detection and recovery in current routed GbE networks.
- an implementation of cost-effective GMPLS shared protection in an IP over 10 GbE testbed.

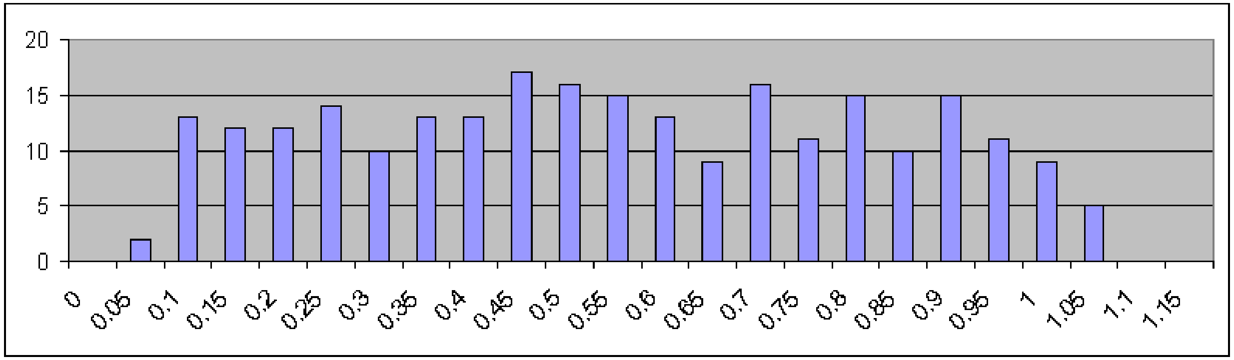
Failure Detection in GbE Point-to-Point network



- PXC is inserted to cause the failure (fiber cut)
- The Linux Box:
  1. Triggers the switching of the PXC thus determining the failure
  2. Detects the LOL through the control circuit
  3. Continuously receives from the Router the XML messages describing the stutus (*up/down*) of the GbE interface

Failure detection time

- *LOL-Status DOWN* delay distribution experimented by the router:



Time [s]

Average delay: 0.573 s

Min value: 92 ms

Max value: 1.091 s.

- The operational state of GbE interfaces is checked by commercial routers just **once** a second

→ mean failure recovery time significantly **high**
(despite of the recovery method e.g., MPLS Protection or Fast Reroute)

OSPF adjacency
in *GbE* Point-to-Point network

- Upon failure detection, each router removes* from its OSPF routing tables the entries referring to the adjacent router and to the networks announced through it.
- Once the connection is physically recovered, it is necessary to wait until the adjacency is re-established and the routing tables synchronized.

Ethernet interfaces are considered Broadcast interfaces
→ time-consuming (~40s)
message exchange to elect the Designated Router (DR) and the Backup DR (BDR).

Interface Operative status UP

OSPF 2-way state

DR and BDR election completed

First data packet after physical activation

No.	Time(sec)	Source	Destination	Protocol	Info
1	0.000000	10.0.30.1	224.0.0.5	OSPF	Hello Packet
2	0.575253	10.0.30.2	224.0.0.5	OSPF	Hello Packet
5	1.580115	10.0.30.1	224.0.0.5	OSPF	Hello Packet
7	9.140728	10.0.30.1	224.0.0.5	OSPF	Hello Packet
8	10.466128	10.0.30.2	224.0.0.5	OSPF	Hello Packet
9	18.191481	10.0.30.1	224.0.0.5	OSPF	Hello Packet
10	18.466769	10.0.30.2	224.0.0.5	OSPF	Hello Packet
11	27.642206	10.0.30.1	224.0.0.5	OSPF	Hello Packet
12	27.947534	10.0.30.2	224.0.0.5	OSPF	Hello Packet
13	35.262820	10.0.30.1	224.0.0.5	OSPF	Hello Packet
14	36.818328	10.0.30.2	224.0.0.5	OSPF	Hello Packet
15	39.043118	10.0.30.1	10.0.30.2	OSPF	DB Descr.
16	39.618474	10.0.30.2	10.0.30.1	OSPF	DB Descr.
17	39.653140	10.0.30.1	10.0.30.2	OSPF	DB Descr.
18	39.688430	10.0.30.2	10.0.30.1	OSPF	LS Request
19	39.723254	10.0.30.1	10.0.30.2	OSPF	LS Update
20	39.723772	10.0.30.2	10.0.30.1	OSPF	DB Descr.
21	39.763110	10.0.30.1	10.0.30.2	OSPF	LS Request
22	39.798439	10.0.30.2	10.0.30.1	OSPF	LS Update
23	39.798912	10.0.30.1	10.0.30.2	OSPF	DB Descr.
24	40.013778	10.0.30.1	224.0.0.5	OSPF	Hello Packet
25	40.588533	10.0.30.2	224.0.0.5	OSPF	Hello Packet
26	40.758537	10.0.30.2	224.0.0.5	OSPF	LS Ack.
27	40.843313	10.0.30.1	224.0.0.5	OSPF	LS Ack.
28	41.593269	10.0.30.1	224.0.0.5	OSPF	Hello Packet
29	42.598732	10.0.30.2	224.0.0.5	OSPF	Hello Packet
30	44.073577	10.0.30.1	224.0.0.5	OSPF	LS Update
31	44.408851	10.0.30.2	10.0.30.1	OSPF	LS Request
32	44.443573	10.0.30.1	10.0.30.2	OSPF	LS Update
33	44.458853	10.0.30.2	224.0.0.5	OSPF	LS Update
34	44.648855	10.0.30.2	224.0.0.5	OSPF	LS Update
35	44.719433	10.0.82.3	10.0.80.3	TCP	
435	45.118942	10.0.30.2	224.0.0.5	OSPF	LS Ack.
821	45.503613	10.0.30.1	224.0.0.5	OSPF	LS Ack.

OSPF adjacency in *ATM* Point-to-Point network

- ATM interfaces can be declared as Point-to-Point interfaces
  → no need to elect the Designated Router (DR) and the Backup DR (BDR).

Interface Operative status UP

First data packet after physical activation

No.	Time(sec)	Source	Destination	Protocol	Packet Type
1	0.000000	10.0.50.1	224.0.0.5	OSPF	Hello Packet
2	1.300000	10.0.50.1	224.0.0.5	OSPF	DB Descr.
3	1.383334	10.0.50.1	224.0.0.5	OSPF	DB Descr.
4	1.433334	10.0.50.1	224.0.0.5	OSPF	LS Request
5	1.466667	10.0.50.1	224.0.0.5	OSPF	DB Descr.
6	1.500000	10.0.50.1	224.0.0.5	OSPF	DB Descr.
7	1.583334	10.0.50.1	224.0.0.5	OSPF	LS Update
8	2.283334	10.0.50.1	224.0.0.5	OSPF	Hello Packet
9	2.483334	10.0.50.1	224.0.0.5	OSPF	LS Acknowledge
10	4.066667	10.0.50.1	224.0.0.5	OSPF	LS Update
11	4.166667	10.0.80.2	200.200.1.1	TCP	

→ A simple router configuration statement (as available for ATM int.) could be introduced for GbE int. to avoid the default Broadcast procedure

(No modifications are required to the OSPF protocol)

GMPLS shared protection scheme

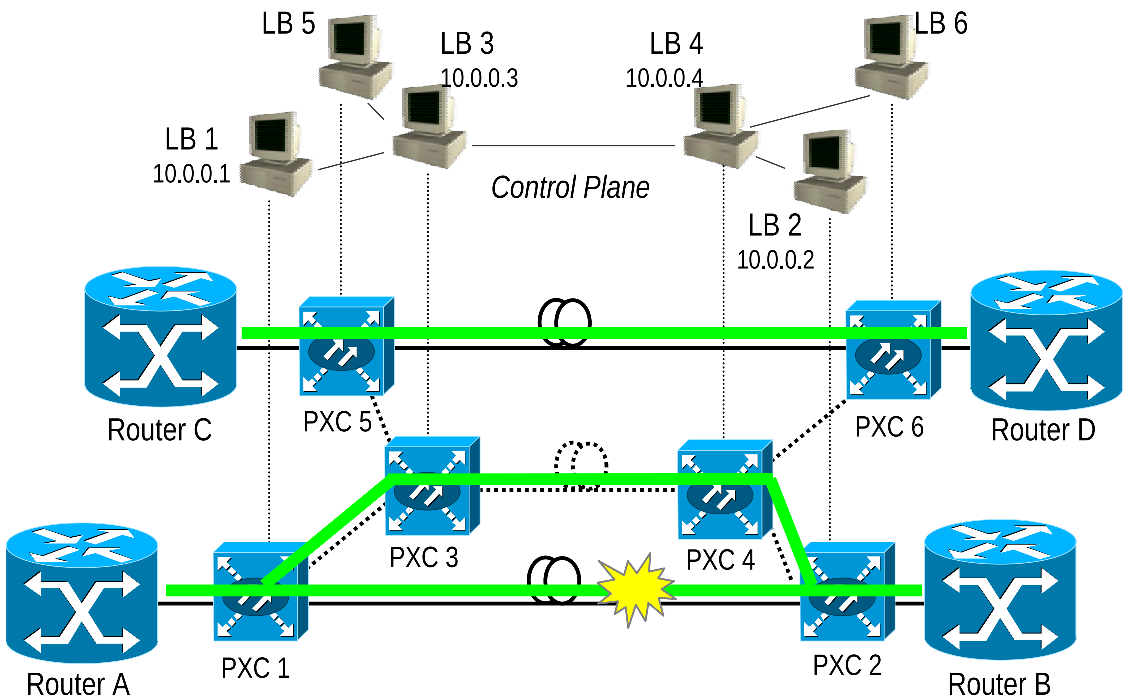
- GMPLS can be used to take advantage of all-optical Network Elements, e.g. transparent Photonic Cross-Connects (PXC).

- This makes possible the realization of shared protection scheme, using shared fibers as backup paths, thus avoiding:
  - the use of expensive electro-optical conversion devices
  - the duplication of GbE interfaces in IP/MPLS routers
  - the duplication of fibers
  - the previously described limitations that affect current IP/MPLS routers.

- We realized the distributed out-of-band control plane to control PXCs.
  At this purpose, some features of 2 protocols:
  - ✓ Link Management Protocol (**LMP**)
  - ✓ Reservation Protocol with GMPLS Extensions (**RSVP-GMPLS**)

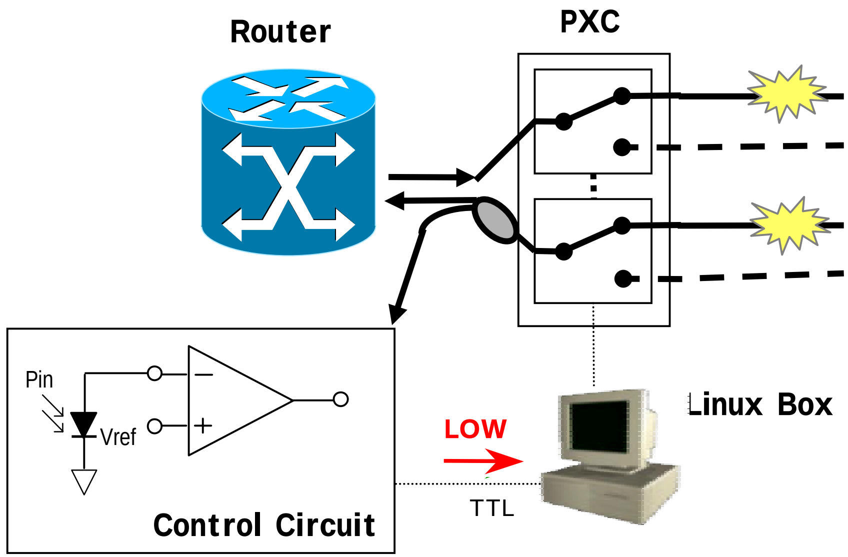
  have been implemented on Linux Box (LB) using C code.

GMPLS shared protection implementation



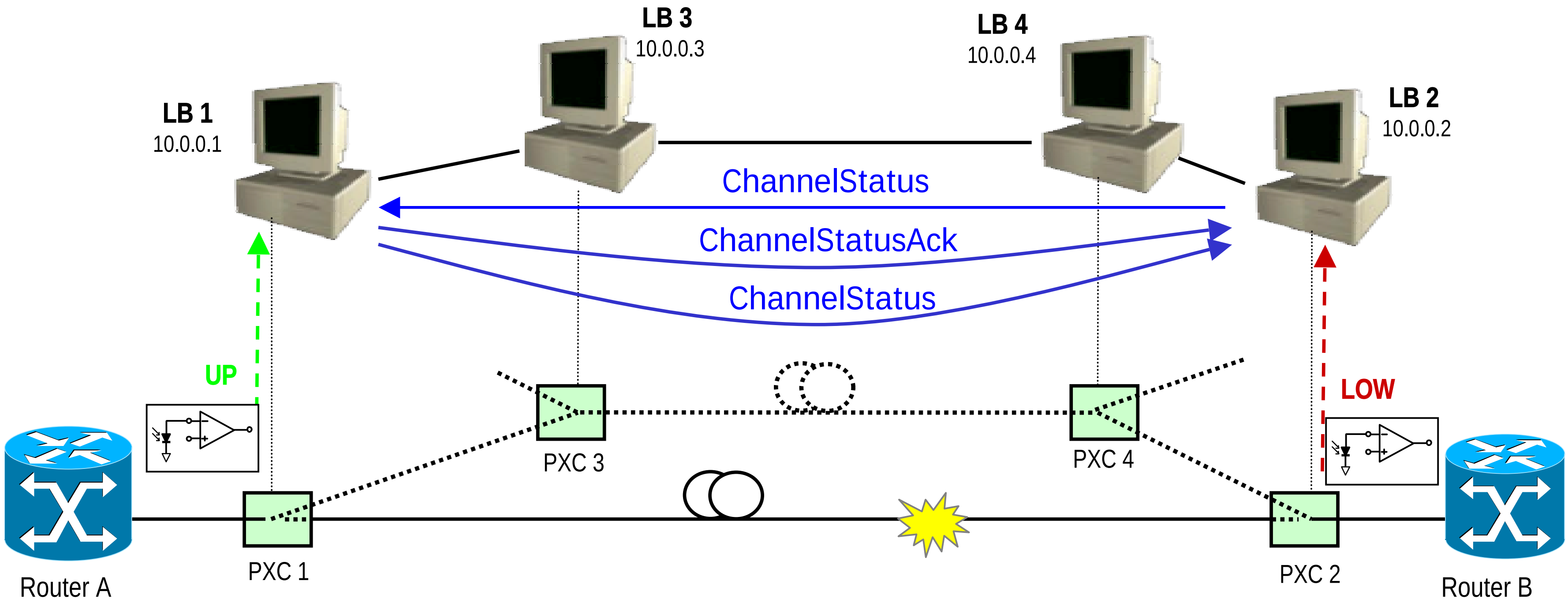
- Initially the primary fibers are used
- If a failure occurs the shared backup fibers are used

Failure Detection mechanism based on Loss of Light (LoL)



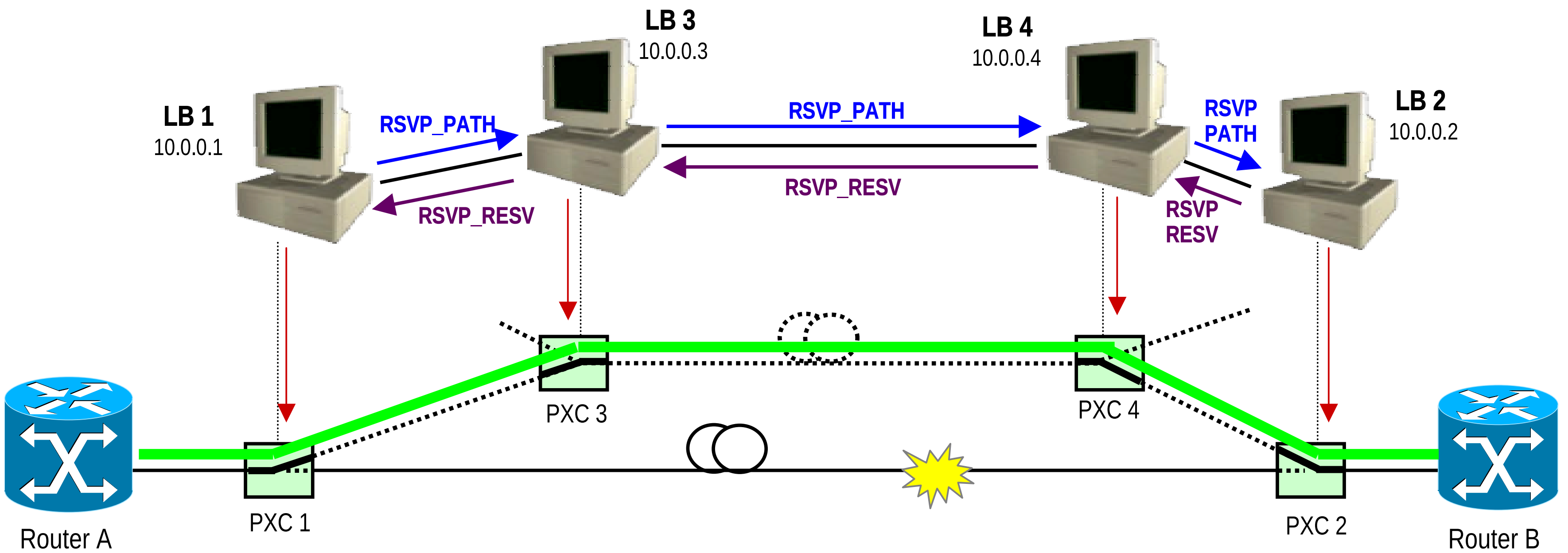
- The received optical signal is <u>split</u> in two fibers
  → part of the signal enters in the control circuit.
- When a failure occurs the output of the Control Circuit becomes *LOW*

Failure Localization
LMP message exchange



- The control circuit detects the Loss of Light (LoL).
- LB2 localizes the failure with its upstream LB1 by exchanging Link Management Protocol (LMP) messages.

Recovery using pre-calculated route



- The upstream LB1 starts the recovery procedure using a pre-calculated path, sending a RSVP PATH message.
- After receiving the proper message each LB emits the switch command and propagates the RSVP RESV message

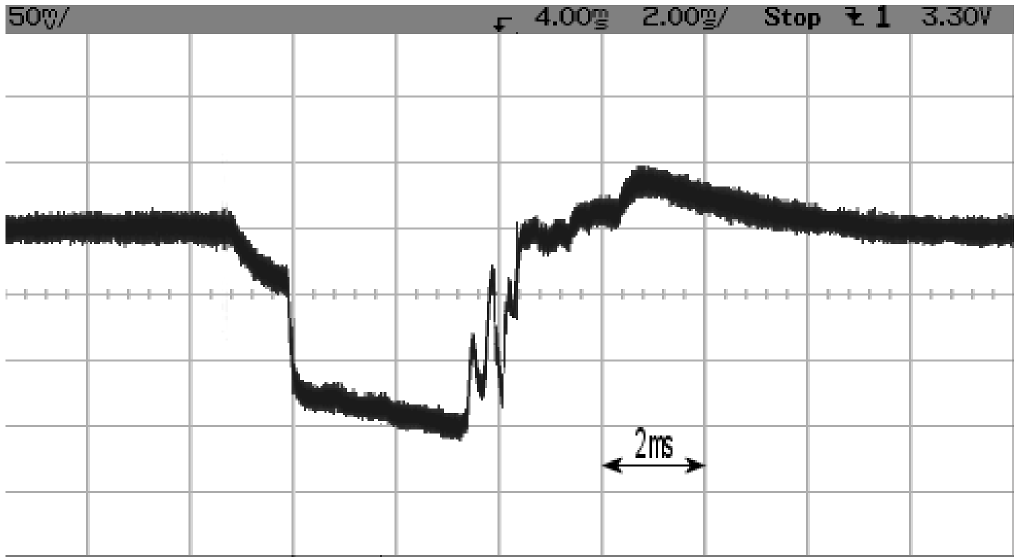
Performance

- The control protocols were implemented on Linux Box in C code

No.	Time	Source	Destination	Protocol	Info
1	0.000000	10.0.0.2	10.0.0.1	LMP	ChannelStatus Message.
2	0.000338	10.0.0.1	10.0.0.2	LMP	ChannelStatusAck Message.
3	0.000465	10.0.0.1	10.0.0.2	LMP	ChannelStatus Message.
4	0.000597	10.0.0.1	10.0.0.3	RSVP	PATH Message. SESSION: IPv4
5	0.000793	10.0.0.3	10.0.0.4	RSVP	PATH Message. SESSION: IPv4
6	0.001107	10.0.0.4	10.0.0.2	RSVP	PATH Message. SESSION: IPv4

- The overall packet exchange takes less than 2 ms.

Performance (2)



- An outage time of 5 ms has been observed
- The speed of this solution is limited almost only by the switching time of the switches, which is less than 5 ms

Conclusions

- This study has experimentally shown two significant limitations that affect routed GbE point-to-point connections between current commercial IP/MPLS routers. Limitations refers to:
  - the delay introduced by the router to detect the failure
  - the time-consuming procedure employed to re-established the routing adjacency upon the physical connectivity is restored.

- Moreover this study has shown that the utilization of GMPLS distributed control plane combined with low cost all-optical network elements allows the cost effective implementation of fast shared protection schemes which avoid the aforementioned limitations and the duplication of resources.

- A recovery time of less than 5 ms has been achieved.